Expanding Access to Tutoring: A Scalable Platform for Personalized Learning and Data-Driven Research

Salome Aguilar Llanes[1,2][0009−0009−7478−8965], Bernardo Garcia Bulle Bueno[1,2][0000−0002−8660−7123], Maria Fernanda Albo Alarcón[2][0009−0004−7337−0800], and Sebastián Guevara Cota[2][0009−0007−1412−8231]

[1] Massachusetts Institute of Technology, USA
[2] Jóvenes Ayudando a Niñas y Niños AC, Mexico.

Abstract. Expanding access to high-quality tutoring is critical for reducing educational disparities, yet scaling effective programs remains a challenge. We developed a platform that automates key logistical aspects of online tutoring, enabling large-scale implementation. Our system includes a real-time monitoring framework that tracks tutor activities. Prior research on online tutoring has shown positive effects on student learning. Building on this, we conducted a randomized controlled trial (RCT) in Mexico. We find that students assigned to tutoring improved their math scores by 0.14 standard deviations. Beyond tutoring delivery, the platform serves as a tool for research. Participating tutors upload class recordings. While this paper focuses on the tutoring intervention, we provide an overview of the platform's potential to facilitate large-scale RCTs. We also show some basic applications of machine learning tools to our data with the aim to analyze student-tutor interactions at scale, bridging the gap between quantitative and qualitative research in education.

Keywords: Online tutoring · Pedagogy · Effects on learner

1 Introduction

Mathematics education is fundamental in shaping students' academic and professional trajectories, improving critical thinking and problem-solving skills.

However, access to quality mathematics education remains limited in many regions, including Latin America, where educational inequalities persist. In Mexico, for example, the 2018 PISA assessment found that more than a third (35%) of students did not reach basic proficiency (level 2) in math, far exceeding the OECD average of 13% [12]. The COVID-19 pandemic further exacerbated this crisis, with estimated learning losses in math of 0.62 to 0.82 standard deviations [8].

One promising intervention to mitigate learning gaps is tutoring; research has demonstrated its efficacy in improving student learning outcomes in various contexts [4][2][11]. Nevertheless, scaling effective tutoring programs remains a logistical and financial challenge, particularly in low-resource settings. Few large-scale tutoring initiatives integrate rigorous impact evaluation frameworks.

Multiple initiatives in the AIED community, such as Cognitive Tutor [14] and ASSISTments [13], have aimed to expand access to math tutoring. However, these solutions rely on high-tech infrastructure, requiring computer access and self-regulated learning skills, restricting their applicability. As a result, many studies focus on small-scale implementations with low technology usage, leaving the intersection of low-resource tutoring and large-scale educational research largely unexplored.

To address these gaps, we developed a scalable, technology-enabled tutoring platform, first deployed in Mexico in 2021 with the mission of providing free, online math tutoring to children. Our platform is a Learning Management System (LMS) specifically designed for tutoring delivery. Past research has examined factors influencing LMS adoption and user satisfaction [6, 1], and more recently has explored the integration of artificial intelligence (AI) and learning analytics to enable adaptive feedback and personalized learning pathways [9]. Our project focuses on using an LMS to implement e-learning through an online tutoring program. Rather than evaluating LMS design or AI integration, we assess the impact of receiving tutoring through the platform versus not receiving it. We also make use of AI tools for research purposes.

The platform matches university-level tutors with small groups of children. We ran a longitudinal study of 15-weeks where 349 tutors were matched to 1910 students (aged 8 to 12) and conducted weekly tutoring sessions. Through a randomized controlled trial (RCT) we assess the effectiveness of the NGO's tutoring program in improving math performance. Results from the large-scale RCT show that students who received tutoring improved their math scores by 0.14 standard deviations.

Finally, we explore how this platform and the data it generates, more than 300,000 of instructional hours recorded, can be leveraged for education research, using AI tools to analyze tutoring interactions and presenting some basic summary statistics.

2 Description of tool

Our platform automates the logistics of large-scale online tutoring and implementing research-driven interventions. On the platform, participants provide informed consent to take part in research. Students and their parents register and complete a baseline assessment, which includes a mathematics diagnostic and a demographic survey. The system then assigns students to groups of one to five peers, based on configurable criteria such as baseline math performance, gender, or socioeconomic background. For this study, groups were either randomly assigned or grouped by baseline math performance [3]. Students were grouped with

[3] We found no meaningful differences between the two assignment methods. These results are not presented here as this is not the primary focus of this analysis.

classmates from their own school. Each group was then matched with a university student tutor, who initiated contact via WhatsApp. Over a defined period, students received two hours of online tutoring per week, followed by an endline assessment to evaluate learning gains and other outcomes.

Tutors register and complete onboarding, which includes asynchronous training modules (covering relevant pedagogical practices, math content, and orientation to the program's logistics) and a group interview. They are then assigned groups of students and begin tutoring. After each session, tutors report attendance, they also detail the material they covered, and student progress. They are required to upload audio recordings of the sessions.

To support education research, the platform enables controlled studies on tutoring effectiveness by allowing researchers to modify student group composition, pair students with tutors based on specific criteria, or randomize access to different instructional strategies.

3 Experiment and NGO in Mexico

In Mexico, university students must complete 480 hours of community service to receive their degree. We launched this tool through an initiative that has since evolved into an established NGO: Jóvenes Ayudando a Niñas y Niños AC (www.jann.mx) and provided, to date, more than 24,500 students with free online tutoring. The NGO collaborates with universities to recruit tutors and works with state education departments to promote tutoring in schools.

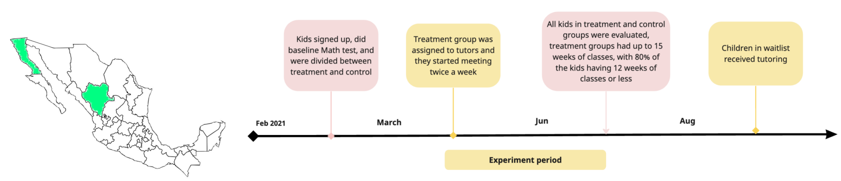

Fig. 1: Left: Location of the states in Mexico where the RCT was performed. Right: Timeline of the experiment.

To measure the impact, in 2021 we provided free tutoring to 3rd and 6th grade students in the states of Durango and Baja California. Students were assessed using a math performance test designed by us based on the national standarized tests PLANEA and ENLACE [5]. They were then assigned to either a control or treatment group through a waitlist design, where the students in the control group experienced a delayed intervention, receiving tutoring in the following academic year. Demographic variables were comparable across treatment and control groups, with no significant differences, except for slightly higher Wifi access in the control group. Details of the timeline are shown in Figure 1. To assess the effect of tutoring, we compared the performance of students in the treatment and control groups. Around 65% of students completed the survey, with similar attrition across groups. To measure the effect of tutoring we ran the following specification:

$$y_{i,1} = \alpha + \beta_1 * t_i + \beta_2 * y_{i,0} + X_i + \epsilon$$

Where $y_{i,1}$ is the endline performance of student i, t_i indicates whether the student was treated, $y_{i,0}$ is the baseline performance on the math test for student i in the baseline, and X_i are other demographic variables.

Dependent variable	Endline performance	Endline performance	Endline performance
Constant	0.015	0.018	-0.02
	(0.066)	(0.06)	(0.115)
Received tutoring	0.127*	0.142**	0.165**
	(0.072)	(0.065)	(0.069)
Baseline performance		0.319***	0.317***
		(0.025)	(0.027)
Family member in university			0.127**
			(0.054)
Demographic controls	No	No	Yes
rsq	0.003	0.144	0.155
N	1245	1245	1157

Table 1: Shows the results of running the specification outlined above. The first column is a regression which only includes the first two terms α,β_1, the second includes β_2 and the third column includes all variables. That column, which includes demographic controls, covers all balance test variables: gender, computer ownership, student with own private room, access to Wifi, and whether a family member attended university. Only the latter yielded significant results, so the other variables were omitted from the table.

On table 1 we observe the coefficients for being assigned treatment in the experiment on the standardized endline math test performance. Regardless of the controls used, there is some significance (between 10 and 5% p-values); suggesting that on average, being assigned the treatment of 2 months of tutoring had an impact of about 0.14 standard deviations on math learning.

4 Recordings

The NGO has compiled over 300,000 hours of class recordings. We save the Mel-Frequency Cepstral Coefficients (MFCC) and transcribe the audios using WhisperX [3]. This rich dataset captures diverse student-tutor dynamics and provides an opportunity to study how these evolve over time.

For this analysis, we focus on recordings from the 2022-2023 school year, which constitute a different sample than the one presented in the previous RCT section.[4]

[4] This is because the recording infrastructure was not fully implemented in 2021 when the RCT was conducted. By 2022, the NGO had expanded to include middle school students, so the sample consists of children ranging in age from 8 to 15. Students were continuously assigned to tutoring groups on a weekly basis throughout the school year.

Our analysis presents summary statistics on the evolution of classroom dynamics both within individual sessions and across the tutoring period. We examine the frequency of exclamation marks in the first five minutes of class transcriptions [5], using them as a proxy for emotional expressiveness and engagement. Prior research [7] has found limitations in the accuracy of punctuation. The main aim of this exercise is to illustrate the richness of the data through one possible use case, rather than to overemphasize the role or reliance of exclamation marks.

Shown in Figure 2, by quantifying the occurrence of exclamation marks in transcribed dialogues, we observe different patterns between primary and middle school students, with exclamation marks appearing more frequently in primary school students. Additionally, as shown in Figure 3 when segmenting the data by tutor gender, classes led by female tutors exhibit a higher prevalence of exclamation marks in their transcriptions.

Focusing on the first class (Figure 4), the divergence between middle and primary school students becomes more pronounced. About 30 minutes in, primary school students show an increase in exclamation mark usage. Similarly other trends can be found when processing the transcriptions to generate Bidirectional Encoder Representations from Transformers (BERT) [10] embeddings and extracting the first two principal components[6].

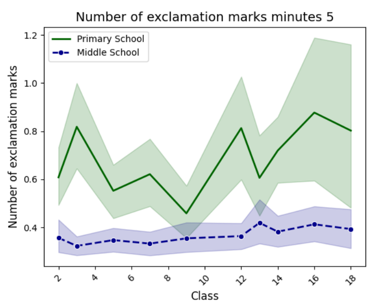

Fig. 2: By school grade

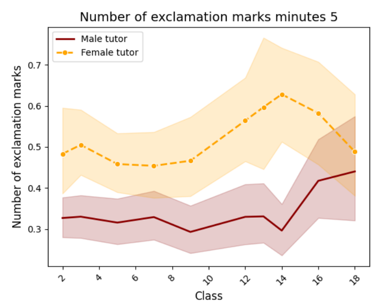

Fig. 3: By gender of tutor

These figures show the mean and standard deviation of exclamation marks per class, for the first 5 minutes of class, disaggregated by the grade and gender of the tutor, across the tutoring period. Each point represents the average number of exclamation marks in a given class session, with error bars indicating the standard deviation.

5 Discussion

We presented a tool, an LMS developed by us, designed to implement large-scale tutoring interventions. We report the results of a randomized controlled trial conducted in Mexico that connected tutors with students to teach math,

[5] Results are similar for minutes 10 to 15

[6] Results not included in this paper.

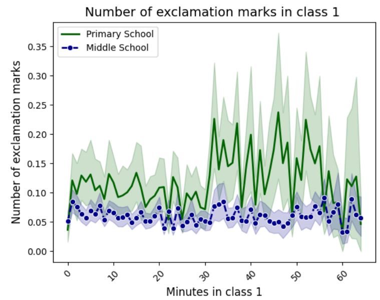

Fig. 4: First class for all groups in the sample by grade.
This figure shows the mean and standard deviation of exclamation marks per minute, disaggregated by the grade, across the first 60 minutes of the first class. Each point represents the average number of exclamation marks in a given class session, with error bars indicating the standard deviation.

showing a significant positive effect on learning, in line with evidence from other remote tutoring interventions [4][2]. Operating in a Latin American context with unpaid university student volunteers, the program highlights the feasibility of low-cost, scalable models.

We then explore the dataset of classroom recordings. Using the first five minutes of each session, we present summary statistics and illustrate how simple linguistic features such as the frequency of exclamation marks can reflect classroom dynamics. This is just one of many analyzable features, others could include silence, number of questions, total word count, sentiment, vocabulary richness, and shifts between instructional and punitive language [15], etc.

Among the limitations of the study are, first, the analysis of the classroom recordings is preliminary and largely descriptive. For example, we use exclamation marks frequency as a proxy for emotional expressiveness, but punctuation restoration in automated transcripts, especially in non-English languages, can be unreliable [7], calling for caution in interpretation. Second, the intervention experienced an attrition rate of 35%. As with many fully online programs, following up with all participants proved challenging. Third, despite being designed for low-resource settings, the intervention still requires access to a mobile device and a stable internet connection, these barriers remain significant for some children.

Nonetheless, the analyses presented here offer a foundation for future research on how participation patterns evolve. The tool enables researchers to investigate questions such as how tutor–student rapport develops and how this, in turn, influences learning outcomes. These insights can inform the design of large-scale, data-driven tutoring interventions. In fact, our team is currently implementing a white-label version of this platform (Intendere) in Peru to test its scalability. The goal of Intendere (intendere.app) is to make our tool accessible to other institutions, allowing them to run similar large-scale tutoring programs and, if desired, conduct research or A/B testing within their own contexts.

Acknowledgments.
This project was supported by the George and Obie Shultz Fund at MIT, which contributed to expenses related to the rollout of the randomized controlled trial and recordings collection; the J-WEL Grant in pK-12 Education Innovation 2022; and the Weiss Fund for Research in Development Economics, which provided additional support for the recordings data collection. We are grateful to Prof. Cimenna Chao at Universidad Iberoamericana for developing the pedagogical content used to train tutors in the program. We also thank the implementation team: Jimena Aguilar Curiel, Carlos Aburto, Maria Maus, Gabriela Curiel, Sebastián Aguilar, and Wenceslao Canett Higuera, for their essential contributions to the field operations. IRB MIT COUHES protocol 2101000304 AEA RCT Registry https://doi.org/10.1257/rct.7307-2.0
Disclosure of Interests.The authors declare no competing interests.

References

1. Al-Fraihat, D., Joy, M., Masa'deh, R., Sinclair, J.: Evaluating e-learning systems success: An empirical study. Computers in human behavior **102**, 67–86 (2020)
2. Angrist, N., Bergman, P., Matsheng, M.: School's out: Experimental evidence on limiting learning loss using "low-tech" in a pandemic. Working Paper 28205, National Bureau of Economic Research (December 2020). https://doi.org/10.3386/w28205, http://www.nber.org/papers/w28205
3. Bain, M., Huh, J., Han, T., Zisserman, A.: Whisperx: Time-accurate speech transcription of long-form audio. arXiv preprint arXiv:2303.00747 (2023)
4. Carlana, M., La Ferrara, E.: Apart but connected: Online tutoring and student outcomes during the covid-19 pandemic. Tech. Rep. RWP21-001, HKS Faculty Research Working Paper Series (February 2021)
5. para la Evaluación de la Educación, I.N.: Base de datos planea - aplicación a tercer grado de preescolar del ciclo escolar 2017-2018 (2018), https://www.inee.edu.mx/evaluaciones/planea/base-de-datos-planea-aplicacion-a-tercer-grado-de-preescolar-del-ciclo-escolar-2017-2018/, available on February 19 2025
6. Gorshenin, A.: Toward modern educational it-ecosystems: from learning management systems to digital platforms. In: 2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). pp. 1–5 (2018). https://doi.org/10.1109/ICUMT.2018.8631229
7. Gris, L.R.S., Marcacini, R., Junior, A.C., Casanova, E., Soares, A., Aluísio, S.M.: Evaluating openai's whisper asr for punctuation prediction and topic modeling of life histories of the museum of the person. arXiv preprint arXiv:2305.14580 (2023)
8. Hevia, F.J., Vergara-Lope, S., Velásquez-Durán, A., Calderón, D.: Estimation of the fundamental learning loss and learning poverty related to covid-19 pandemic in mexico. International Journal of Educational Development **88**, 102515 (2022). https://doi.org/10.1016/j.ijedudev.2021.102515, https://doi.org/10.1016/j.ijedudev.2021.102515
9. Ifenthaler, D., Yau, J.Y.K.: Utilising learning analytics to support study success in higher education: a systematic review. Educational Technology Research and Development **68**(4), 1961–1990 (2020)
10. Kenton, J.D.M.W.C., Toutanova, L.K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of naacL-HLT. vol. 1. Minneapolis, Minnesota (2019)

11. Nickow, A., Oreopoulos, P., Quan, V.: The promise of tutoring for prek–12 learning: A systematic review and meta-analysis of the experimental evidence. American Educational Research Journal **61**(1), 74–107 (2024)
12. OECD: Mexico country note. OECD Publishing (2018), https://doi.org/10.1787/eag-2018-en
13. Pane, J.F., Griffin, B.A., McCaffrey, D.F., Karam, R.: Effectiveness of cognitive tutor algebra i at scale. Educational Evaluation and Policy Analysis **36**(2), 127–144 (2014)
14. Ritter, S., Anderson, J.R., Koedinger, K.R., Corbett, A.: Cognitive tutor: Applied research in mathematics education. Psychonomic bulletin & review **14**, 249–255 (2007)
15. Tan, M., Demszky, D.: Sit down now: How teachers' language reveals the dynamics of classroom management practices (2023)